

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ імені ІГОРЯ СІКОРСЬКОГО»

БРАТУСЬ ОЛЕНА ВІКТОРІВНА

УДК 338.27, 681.5.015

**МЕТОДИ ПРОГНОЗУВАННЯ НЕСТАЦІОНАРНИХ ЧАСОВИХ РЯДІВ НА
ОСНОВІ ДВОБІЧНОГО ЕКСПОНЕНЦІЙНОГО ЗГЛАДЖУВАННЯ ТА
ОПТИМАЛЬНОЇ ФІЛЬТРАЦІЇ**

01.05.04 – системний аналіз і теорія оптимальних рішень

АВТОРЕФЕРАТ
дисертації на здобуття наукового ступеня
кандидата технічних наук

Київ – 2019

Дисертацією є рукопис.

Робота виконана на кафедрі математичних методів системного аналізу Інституту прикладного системного аналізу Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського» Міністерства освіти і науки України.

Науковий керівник: доктор технічних наук, професор
Бідюк Петро Іванович,
Інститут прикладного системного аналізу
Національного технічного університету України
«Київський політехнічний інститут імені
Ігоря Сікорського», професор кафедри
математичних методів системного аналізу.

Офіційні опоненти: доктор технічних наук, професор
Яценко Віталій Олексійович,
Інститут космічних досліджень НАН України та
Державного космічного агентства України,
завідувач відділу дистанційних методів та перспективних
приладів;

доктор технічних наук, професор
Корабльов Микола Михайлович,
Харківський національний університет радіоелектроніки,
професор кафедри електронних обчислювальних машин.

Захист відбудеться «22» жовтня 2019 р. о 17⁰⁰ годині на засіданні спеціалізованої вченої ради Д 26.002.03 при Національному технічному університеті України «Київський політехнічний інститут імені Ігоря Сікорського» за адресою: 03056, м. Київ, пр. Перемоги, 37, корп. 35, ауд. 001.

З дисертацією можна ознайомитись у бібліотеці Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського» за адресою: 03056, м. Київ, пр. Перемоги, 37.

Автореферат розісланий «__» вересня 2019 р.

Вчений секретар
спеціалізованої вченої ради Д 26.002.03

В.О. Капустян

ЗАГАЛЬНА ХАРАКТЕРИСТИКА РОБОТИ

Актуальність теми. Дисертаційна робота присвячена розробці та застосуванню прикладної наукової методології системного аналізу для розв'язання задач відновлення пропущених значень часових рядів, відновлення істинних закономірностей розвитку досліджуваних процесів, прогнозування часових рядів та виконання автоматизованої обробки часових рядів за допомогою створеної системи підтримки прийняття рішень. Відновлення пропущених значень часових рядів є важливою проблемою на сьогоднішній день, тому що це дає змогу правильно виконати попередню обробку часових рядів та їх подальше прогнозування. Відновлення істинних закономірностей досліджуваних процесів є актуальним для часових рядів, що містять значні шумові складові та потребують попередньої обробки перед побудовою математичної моделі та прогнозуванням. Отримання точніших оцінок значень часових рядів є важливим для тих, хто займається прогнозуванням часових рядів у своїй повсякденній роботі, наприклад, займається купівлею та продажем металів на біржі металів або інших активів. Розробка нових систем підтримки прийняття рішень для аналізу та прогнозування часових рядів на основі сучасних принципів системного аналізу є важливою задачею, оскільки такі вдосконалені системи дадуть змогу виконувати швидко та більш точну обробку даних.

Важливі теоретичні дослідження у галузі системного аналізу виконані багатьма видатними вченими: академіком В.І. Вернадським, О.О. Богдановим, Л. фон Берталанфі, Т. Котарбінським, Н. Вінером. Ці дослідження були у подальшому розвинуті у сучасних роботах академіка М.З. Згуровського, член-кореспондента Н.Д. Панкратової, В.Д. Романенка, В.Я. Данилова, В.О. Капустяна, О.А. Павлова, Т. Сааті, О.В. Антонова, Н.З. Шора, Р.Л. Акоффа та інших вчених.

Відомими сучасними вченими у галузі аналізу та прогнозування часових рядів є Ю.П. Зайченко, В.Д. Романенко, К.Т. Леондес, Дж. Бендат, А. Пірсол. Докладні дослідження стосовно фільтра Калмана виконано у роботах Р. Мехра, В.М. Подладчікова. Важливою задачею є також ідентифікація параметрів систем для подальшого їх використання при прогнозуванні. Дослідження стосовно методів ідентифікації параметрів систем виконані у роботах К.Й. Острема та П. Ейкоффа. Е.С. Гарднер молодший виконав дослідження стосовно експоненційного згладжування та створив класифікацію методів експоненційного згладжування. Дослідженням відновлення істинних закономірностей процесів та прогнозуванням сонячних даних займалися багато науковців. У роботах Д.Г. Хетеуея, Р.М. Вілсона та Е.Дж. Рейхмана запропоновано рівняння для опису моделі 11-річного сонячного циклу, яке відповідно до їх досліджень може бути використаним для прогнозування. У подальших роботах цих авторів запропоновано метод розрахунку 13-місячного ковзного середнього для обробки даних про сонячну активність, а також виконано докладний огляд багатьох існуючих методів прогнозування сонячних даних.

Існує значна кількість нерозв'язаних задач, що стосуються аналізу та прогнозування часових рядів та є актуальними на сьогоднішній день: підвищення

адекватності математичних моделей часових рядів; отримання точніших оцінок значень часових рядів під час відновлення пропущених значень, відновлення істинних закономірностей та прогнозування часових рядів. Дисертаційна робота присвячена виконанню досліджень стосовно розв'язання актуальних наведених задач, а також розробці оригінального програмного продукту у формі системи підтримки прийняття рішень для забезпечення автоматизованого розв'язання цих задач з використанням традиційних та розроблених у роботі методів.

Зв'язок роботи з науковими програмами, планами, темами. Дисертаційна робота виконана на кафедрі математичних методів системного аналізу Інституту прикладного системного аналізу Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського» у відповідності до планів науково-дослідних робіт:

1. д/б НДР «Розробка інформаційної технології моделювання та оцінювання фінансово-економічних ризиків із врахуванням невизначеностей різної природи (на основі байєсівських моделей)» (№ ДР 0113U000650);
2. д/б НДР «Розробка методології системного аналізу, моделювання та оцінювання фінансових ризиків» (№ ДР 0115U000356);
3. д/б НДР «Проектування сучасних систем сервісів на прикладі мобільної медичної системи для мешканців прифронтових селищ в зоні АТО» (№ ДР 0117U002435).

Мета і завдання дослідження. *Мета дослідження* – розробка методів оцінювання параметрів математичних моделей, що змінюються за невідомим законом, та їх застосування для підвищення якості оцінок прогнозів розвитку часових рядів шляхом удосконалення системної методології прогнозування; розробка методів відновлення пропущених значень, істинних закономірностей та прогнозування часових рядів; розробка оригінального програмного продукту у формі системи підтримки прийняття рішень. Для досягнення мети потрібно вирішити такі завдання:

– для математичної моделі, в якій враховуються швидкість та прискорення зміни значень вибірки даних, створити методи оцінювання математичного сподівання прискорення зміни значень вибірки даних з використанням повних та розріджених вибірок нев'язок субоптимального фільтра з пам'яттю два; обрати кращі методи серед створених та застосувати їх для розробки адаптивних фільтрів Калмана;

– для математичної моделі, в якій враховуються швидкість та прискорення зміни значень вибірки даних, створити метод оцінювання математичного сподівання прискорення зміни значень вибірки даних, що змінюється за невідомим законом, з застосуванням процедури експоненційного згладжування до побудованого ряду однозалежних псевдовимірів цього параметру; застосувати створений метод для розробки алгоритмів прогнозування на основі фільтра Калмана;

– розробити метод двобічного експоненційного згладжування для відновлення пропущених значень, істинних закономірностей та прогнозування часових рядів; розробити інтегральний критерій адекватності моделі для вибору оптимального параметра згладжування з метою застосування у методах, які пов'язані з експоненційним згладжуванням;

- розробити методи відновлення пропущених значень та прогнозування для взаємозалежних часових рядів з використанням двобічного експоненційного згладжування;

- розробити метод ковзного двобічного експоненційного згладжування для відновлення пропущених значень, істинних закономірностей та прогнозування часових рядів; розробити критерій близькості для оцінювання моделей під час відновлення істинних закономірностей часових рядів;

- розробити метод адаптивного ковзного двобічного експоненційного згладжування для відновлення пропущених значень, істинних закономірностей та прогнозування часових рядів;

- створити архітектуру та структурну схему системи підтримки прийняття рішень для організації, що займається аналізом та прогнозуванням часових рядів, а також виконати її програмну реалізацію мовою програмування Java.

Об'єкт дослідження: процеси, що описуються фінансово-економічними часовими рядами, показниками сталого розвитку, сонячними даними (числами Вольфа, сонячними радіофлюксами при довжині хвилі 10,7 см).

Предмет дослідження: моделі і методи відновлення пропущених значень, істинних закономірностей та прогнозування часових рядів.

Методи дослідження. У роботі використано такі методи дослідження: методи регресійного аналізу, адаптивний регресійний підхід, відновлення пропущених значень часових рядів за допомогою середнього арифметичного, метод групового урахування аргументів, методи оцінювання математичного сподівання прискорення зміни значень вибірки даних, адаптивні фільтри Калмана, алгоритми прогнозування на основі фільтра Калмана, метод експоненційного згладжування, метод подвійного експоненційного згладжування, експоненційний підхід до опису сонячного циклу, метод 13-місячного ковзного середнього, метод двобічного експоненційного згладжування, метод ковзного двобічного експоненційного згладжування, метод адаптивного ковзного двобічного експоненційного згладжування, методи відновлення пропущених значень та прогнозування для взаємозалежних часових рядів з використанням двобічного експоненційного згладжування.

Наукова новизна одержаних результатів. Виконані у дисертаційній роботі дослідження дозволили отримати такі теоретичні та практичні результати:

Уперше:

- для математичної моделі, в якій враховуються швидкість та прискорення зміни значень вибірки даних, створено методи оцінювання математичного сподівання прискорення зміни значень вибірки даних, що змінюється за невідомим законом, запропоновані методи дають можливість оцінити цей невідомий параметр;

- розроблено метод двобічного експоненційного згладжування для відновлення пропущених значень, істинних закономірностей та прогнозування часових рядів, що дає змогу отримувати точніші оцінки часового ряду за рахунок врахування інформації про часовий ряд, отриманої у зворотному часі; розроблено інтегральний критерій адекватності моделі для вибору оптимального значення параметра згладжування, який відрізняється одночасним врахуванням двох статистик і забезпечує вибір адекватної моделі;

– розроблено методи відновлення пропущених значень та прогнозування для взаємозалежних часових рядів з використанням двобічного експоненційного згладжування, ці методи дають можливість враховувати динаміку взаємозалежних часових рядів для отримання точніших оцінок значень часових рядів;

– розроблено методи ковзного двобічного експоненційного згладжування та адаптивного ковзного двобічного експоненційного згладжування для відновлення пропущених значень, істинних закономірностей та прогнозування часових рядів, ці методи призначені для роботи з сонячними даними і дають можливість враховувати специфіку таких даних.

Удосконалено:

– адаптивні фільтри Калмана та алгоритми прогнозування на основі фільтра Калмана з використанням методів оцінювання математичного сподівання прискорення зміни значень вибірки даних, що змінюється за невідомим законом, які відрізняються врахуванням оцінок невідомого математичного сподівання прискорення зміни значень вибірки даних і дають можливість отримати точніші оцінки прогнозів часових рядів.

Отримав подальший розвиток:

– критерій близькості для оцінювання моделей під час відновлення істинних закономірностей часових рядів.

Практичне значення одержаних результатів

1. У результаті виконання дослідження розроблено та програмно реалізовано цілісну системну методологію для відновлення пропущених значень, відновлення істинних закономірностей та прогнозування часових рядів. Розроблена система підтримки прийняття рішень застосована для аналізу, відновлення пропущених значень, відновлення істинних закономірностей та прогнозування фінансово-економічних та сонячних даних.

2. Результати дисертації у складі моделей та методів впроваджені у Навчально-науковому комплексі «Світовий центр даних з геоінформатики та сталого розвитку» (акт впровадження № 0301/18 від 13.03.2018). Впровадження запропонованої методології дало можливість отримати високоякісні результати відновлення пропущених значень та оцінок прогнозів для взаємозалежних показників сталого розвитку Хорватії (багатовимірні часові ряди).

3. Результати дисертації у складі моделей, методів та програмних засобів були впроваджені у навчальний процес Інституту прикладного системного аналізу Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського» (довідка про впровадження № 48/4040 від 22.03.2018). З використанням запропонованих методів і математичних моделей розроблено лабораторні роботи для таких дисциплін з навчального плану Інституту прикладного системного аналізу: «Аналіз часових рядів», «Прогнозування у фінансово-економічній сфері» та «Аналіз нелінійних часових рядів». Розроблене програмне та алгоритмічне забезпечення використано для створення сучасної інформаційної технології аналізу фінансових даних, яка взята за основу для лабораторних та курсових робіт з навчальної дисципліни «Системи і методи підтримки прийняття рішень».

Особистий внесок здобувача. Всі основні наукові положення та результати, що складають основний зміст роботи та становлять наукову новизну, отримані автором самостійно. У роботі [1], опублікованій у співавторстві, здобувачеві належать створені методи № 4–6 для оцінювання математичного сподівання прискорення зміни значень вибірки даних, використання розробленого методу № 5 для побудови адаптивного фільтра Калмана та подальшого прогнозування цін на цинк. У роботі [2], опублікованій у співавторстві, здобувачеві належать метод оцінювання математичного сподівання прискорення зміни значень вибірки даних та створений алгоритм прогнозування на основі фільтра Калмана з використанням цього методу. У роботі [3], опублікованій у співавторстві, здобувачеві належать розроблені метод двобічного експоненційного згладжування та інтегральний критерій адекватності моделі. У роботі [5], опублікованій у співавторстві, здобувачеві належать метод відновлення пропущених значень для взаємозалежних часових рядів з використанням двобічного експоненційного згладжування та метод прогнозування для взаємозалежних часових рядів з використанням двобічного експоненційного згладжування. У доповіді [7], опублікованій у співавторстві, здобувачеві належать методи оцінювання математичного сподівання прискорення зміни значень вибірки даних та адаптивний фільтр Калмана. У доповіді [8], опублікованій у співавторстві, здобувачеві належить метод двобічного експоненційного згладжування. У доповіді [11], опублікованій у співавторстві, здобувачеві належать метод ковзного двобічного експоненційного згладжування та критерій близькості.

Апробація результатів дисертації. Основні положення дисертаційної роботи були представлені на наукових конференціях та семінарах: 15^й міжнародній науково-технічній конференції SAIT 2013 (Україна, м. Київ, 27-31 травня 2013 р.); XII міжнародній конференції «Контроль і управління в складних системах (КУСС-2014)» (Україна, м. Вінниця, 14-16 жовтня 2014 р.); міжнародній науковій конференції «Інтелектуальні системи прийняття рішень і проблеми обчислювального інтелекту» (Україна, с. Залізний Порт, 24-28 травня 2016 р.); V^й міжнародній конференції «Математика. Інформаційні технології. Освіта» (Україна, м. Луцьк - с. Світязь, 5-7 червня 2016 р.); Одинадцятій міжнародній науково-практичній конференції «Математичне та імітаційне моделювання систем. МОДС 2016» (Україна, с. Жукин, 27 червня - 1 липня 2016 р.); науковому семінарі «Системні дослідження та інформаційні технології» в Інституті прикладного системного аналізу Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського» (Україна, м. Київ, 21 березня 2018 р.).

Публікації. За матеріалами дисертаційного дослідження опубліковано 11 наукових праць: серед них 6 наукових статей у провідних наукових фахових виданнях [1-6] (у тому числі 1 – в іноземному виданні [6]; 1 – в українському виданні, що входить до міжнародної наукометричної бази даних Web of Science [4]; 1 – в українському виданні, що у повному обсязі перевидається англійською мовою під назвою «Journal of Automation and Information Sciences» та входить до

міжнародної наукометричної бази даних Scopus [5]), 5 праць у матеріалах доповідей міжнародних конференцій [7-11].

Структура та обсяг дисертації. Дисертаційна робота складається з анотації, змісту, переліку скорочень, вступу, чотирьох основних розділів, висновків, списку використаних джерел, трьох додатків. Робота викладена на 194 сторінках і містить 161 сторінку основної частини, 34 рисунки, 21 таблицю, список використаних джерел із 89 найменувань на 9 сторінках, 3 додатки на 5 сторінках.

ОСНОВНИЙ ЗМІСТ РОБОТИ

У вступі обґрунтовано актуальність розробки та застосування прикладної наукової методології системного аналізу для розв'язання задач відновлення пропущених значень часових рядів, відновлення істинних закономірностей розвитку досліджуваних процесів, прогнозування часових рядів та виконання автоматизованої обробки часових рядів за допомогою створеної системи підтримки прийняття рішень. Визначено зв'язок з науковими програмами, планами, темами. Також визначено мету, завдання, об'єкт, предмет і методи дослідження, наведено наукову новизну та практичне значення одержаних результатів, висвітлено особистий внесок здобувача, описано апробацію результатів роботи, публікації за результатами роботи, структуру та обсяг дисертаційної роботи.

У першому розділі дисертаційної роботи виконано огляд сучасних методів аналізу, відновлення істинних закономірностей, пропущених значень та прогнозування часових рядів. Наведено застосування системного підходу до аналізу та прогнозування часових рядів. Описано методики перевірки на наявність нестационарності, а саме, тренду і гетероскедастичності. Наведено сезонні ефекти та їх врахування під час аналізу часового ряду. Описано традиційні методи відновлення пропущених значень часових рядів. Виконано огляд методів ідентифікації моделей динаміки нестационарних процесів. Описано методи і моделі для прогнозування часових рядів: регресійні моделі, методи експоненційного згладжування, фільтр Калмана, метод групового урахування аргументів. Виконано аналіз можливостей застосування систем підтримки прийняття рішень для аналізу та прогнозування часових рядів. Виконаний аналіз виявив необхідність проведення подальших досліджень та дозволив сформулювати мету дослідження: розробка методів оцінювання параметрів математичних моделей, що змінюються за невідомим законом, та їх застосування для підвищення якості оцінок прогнозів розвитку часових рядів шляхом удосконалення системної методології прогнозування; розробка методів відновлення пропущених значень, істинних закономірностей та прогнозування часових рядів; розробка оригінального програмного продукту у формі системи підтримки прийняття рішень.

У другому розділі дисертаційної роботи описано застосування системного підходу до аналізу часових рядів та розробки методів їх прогнозування. Розроблено методи оцінювання математичного сподівання прискорення зміни значень вибірки даних з використанням повних та розріджених вибірок нев'язок субоптимального фільтра з пам'яттю два, та виконано порівняння створених методів з існуючими.

Обрано кращі методи серед розроблених, та застосовано їх для розробки адаптивних фільтрів Калмана. Розроблені адаптивні фільтри Калмана під час прогнозування середньодобових цін Лондонської біржі металів на цинк показали перевагу перед методами експоненційного згладжування, авторегресії, авторегресії з ковзним середнім та групового урахування аргументів за значеннями прогнозних характеристик (наводяться значення MAPE – середньої відсоткової абсолютної похибки (англ. Mean absolute percentage error): MAPE=4,1562% для адаптивного фільтра Калмана № 1; MAPE=4,1176% для адаптивного фільтра Калмана № 2).

Розроблено метод оцінювання математичного сподівання прискорення зміни значень вибірки даних для випадку, коли цей параметр змінюється за невідомим законом, з застосуванням процедури експоненційного згладжування до побудованого ряду однозалежних псевдовимірів цього параметру. Створено алгоритми прогнозування на основі фільтра Калмана з використанням цього розробленого методу. Виконано прогнозування середньодобових цін на свинець за даними Лондонської біржі металів за допомогою створених алгоритмів та порівняння з результатами, отриманими за допомогою методів авторегресії, авторегресії з ковзним середнім та групового урахування аргументів. Створені алгоритми прогнозування на основі фільтра Калмана показали перевагу перед усіма методами за значеннями прогнозних характеристик (MAPE=4,3465% для алгоритму прогнозування № 1 на основі фільтра Калмана; MAPE=4,2352% для алгоритму прогнозування № 2 на основі фільтра Калмана).

У третьому розділі дисертаційної роботи наведено розроблений метод двобічного експоненційного згладжування для відновлення пропущених значень, істинних закономірностей та прогнозування часових рядів. У роботі розглядається процедура експоненційного згладжування для часового ряду зі значенням початкового наближення, що дорівнює першому значенню ряду:

$$S_k = \alpha \times y(k) + (1 - \alpha) \times S_{k-1}, \quad (1)$$

де S_k – значення експоненційного середнього в k -й момент часу; α – параметр згладжування; $y(k)$ – значення ряду в k -й момент часу; S_{k-1} – значення експоненційного середнього в $(k-1)$ -й момент часу.

У роботі запропоновано інтегральний критерій адекватності моделі, що описується таким чином:

$$I = \gamma \times |1 - R^2| + \eta \times |2 - DW|, \quad (2)$$

де R^2 – коефіцієнт детермінації; DW – статистика Дарбіна-Уотсона; γ, η – вагові коефіцієнти. У даній роботі вагові коефіцієнти мають такі значення: $\gamma = 1; \eta = 0,5$.

Критерієм вибору оптимального значення параметра згладжування є мінімум значень суми квадратів похибок SSE (англ. Sum squared error) та інтегрального критерію I , що обчислюється за формулою (2).

У роботі запропоновано застосовувати до експоненційних середніх, отриманих за формулою (1), експоненційне згладжування у зворотному часі $k = N-1, \dots, 1$. Кожне уточнене значення S_k^* обчислюється так:

$$S_k^* = \alpha \times S_k + (1 - \alpha) \times S_{k+1}^*, \quad (3)$$

де S_k^* – значення експоненційного середнього у зворотному часі у k -й момент часу, тут значення $k = N - 1, N - 2, \dots, 1$, початкове значення $S_N^* = S_N$.

Критерієм вибору оптимального значення параметра згладжування є мінімум значень суми квадратів похибок SSE та інтегрального критерію I , що обчислюється за формулою (2), але в якості вхідних значень при обчисленні цих характеристик використовуються значення S_k .

Описано застосування створеного методу двобічного експоненційного згладжування для відновлення істинних закономірностей та прогнозування часових рядів. Розроблено алгоритм для відновлення пропущених значень часового ряду з використанням методу двобічного експоненційного згладжування. Метод двобічного експоненційного згладжування показав перевагу перед відновленням за допомогою середнього арифметичного, авторегресійного підходу та методу експоненційного згладжування за значеннями статистичних характеристик під час відновлення середньодобових значень цін на цинк за даними Лондонської біржі металів.

Розроблено метод відновлення пропущених значень для взаємозалежних часових рядів з використанням двобічного експоненційного згладжування. Цей метод використано для відновлення значень взаємозалежних показників сталого розвитку за даними Світового центру даних, та виконано порівняння зі значеннями, отриманими за допомогою методів відновлення середніми арифметичними та експоненційним згладжуванням. Створений метод показав перевагу перед іншими методами за значеннями статистичних характеристик. Розроблено метод прогнозування для взаємозалежних часових рядів з використанням двобічного експоненційного згладжування. Цей метод використано для прогнозування показників сталого розвитку за даними Світового центру даних, та виконано порівняння з результатами, отриманими за методом експоненційного згладжування. Створений метод показав перевагу перед методом експоненційного згладжування за значеннями прогнозних характеристик.

У даній роботі розроблено метод ковзного двобічного експоненційного згладжування. Перша частина процедури ковзного двобічного тринадцятимісячного експоненційного згладжування полягає в отриманні оцінок за допомогою традиційного методу експоненційного згладжування. При цьому початкове наближення обирається рівним першому значенню ряду з відповідного інтервалу на кожному ковзному тринадцятимісячному інтервалі. Ця процедура описується так:

$$S_k = \alpha \times y(k) + (1 - \alpha) \times S_{k-1}, \quad (4)$$

де S_k – значення експоненційного середнього в k -й момент часу; α – коефіцієнт згладжування; $y(k)$ – значення ряду в k -й момент часу; S_{k-1} – значення експоненційного середнього в $(k - 1)$ -й момент часу, періоди часу $k = 1, \dots, 13, k = 2, \dots, 14, \dots, k = N - 12, \dots, N$.

Критерієм вибору оптимального значення параметра згладжування є мінімум значень суми квадратів похибок SSE та інтегрального критерію I , що обчислюється за формулою (2).

Друга частина процедури ковзного двобічного експоненційного згладжування полягає у застосуванні до отриманих оцінок процедури згладжування у зворотному часі на кожному інтервалі $k = N - 1, \dots, N - 12, k = N - 2, \dots, N - 13, \dots, k = 12, \dots, 1$. Кожне уточнене значення S_k^* розраховується так:

$$S_k^* = \alpha \times S_k + (1 - \alpha) \times S_{k+1}^*, \quad (5)$$

де S_k^* – значення експоненційного середнього у зворотному часі в k -й момент часу; α – коефіцієнт згладжування; S_k – значення експоненційного середнього в k -й момент часу; S_{k+1}^* – значення експоненційного середнього у зворотному часі в $(k+1)$ -й момент часу, тут значення $k = N - 1, \dots, N - 12, \dots, k = 12, \dots, 1$, початкові значення $S_N^* = S_N, \dots, S_{13}^* = S_{13}$.

Критерієм вибору оптимального значення параметра згладжування є мінімум значень суми квадратів похибок SSE та інтегрального критерію I , що обчислюється за формулою (2), але в якості вхідних значень при обчисленні цих характеристик використовуються значення S_k .

Після цього в якості оцінки значення вибірки даних використовується отримане значення у середній точці з кожного інтервалу, тобто на інтервалі $k = 1, \dots, 13$ взято значення S_7^* , на інтервалі $k = 2, \dots, 14$ взято значення S_8^* і т. д. Оцінки для перших шести та останніх шести значень вибірки даних взято як відповідні значення експоненційного середнього у зворотному часі на першому та останньому інтервалах оцінювання відповідно, тобто S_1^*, \dots, S_6^* на інтервалі $k = 1, \dots, 13$ та S_{N-5}^*, \dots, S_N^* на інтервалі $k = N - 12, \dots, N$.

Розроблено алгоритм для відновлення пропущених значень ряду з використанням методу ковзного двобічного експоненційного згладжування. Розроблений алгоритм показав перевагу перед методом експоненційного згладжування за значеннями статистичних характеристик під час відновлення даних про місячні значення сонячних радіофлюксів при довжині хвилі 10,7 см (далі використовується назва індекс F10.7) за даними Національного управління океанічних і атмосферних досліджень Сполучених Штатів Америки.

Розроблено критерій близькості для оцінювання моделей під час відновлення істинних закономірностей часових рядів. На основі вибірки даних про місячні значення індексу F10.7 в одиницях виміру SFU (англ. Solar Flux Unit) з січня 2013 року до березня 2016 року (39 значень) відновлюються істинні значення ряду за допомогою методу ковзного двобічного експоненційного згладжування за формулами (4), (5) та методу 13-місячного ковзного середнього, результати показано на рис. 1.

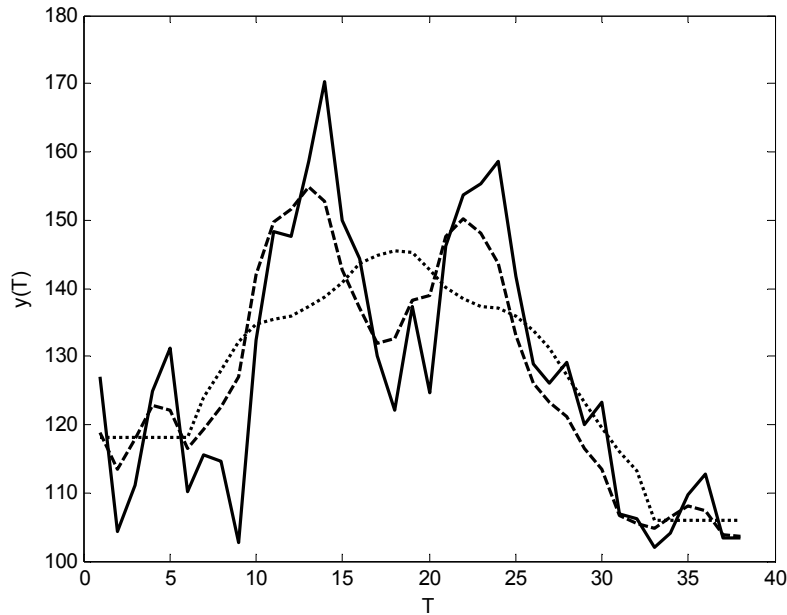


Рис. 1. Значення $y(i)$ та $\hat{y}(i)$ для вибірки значень індексу F10.7, де $-$ – виміряні значення $y(i)$; $--$ – оцінки $\hat{y}(i)$, отримані за методом ковзного двобічного експоненційного згладжування; \cdots – оцінки $\hat{y}(i)$, отримані за методом 13-місячного ковзного середнього

На рис. 1 показано, що на часовому інтервалі $[10, 25]$ для моделі, побудованої за методом 13-місячного ковзного середнього, відбувається суттєве викривлення процесу: на часових інтервалах $[10, 15]$ та $[20, 25]$ коливання обнуляються, на інтервалі $[15, 20]$ відбувається обертання коливань індексу F10.7. На цьому часовому інтервалі метод ковзного двобічного експоненційного згладжування не викривляє коливання під час відновлення істинних закономірностей індексу F10.7. Також цей метод показав перевагу перед методом 13-місячного ковзного середнього при відновленні істинних закономірностей індексу F10.7 за значенням критерію близькості.

За значеннями прогнозних характеристик метод ковзного двобічного експоненційного згладжування показав перевагу перед традиційним експоненційним підходом при прогнозуванні значень сонячних радіофлюксів при довжині хвилі 10,7 см.

Розроблено метод адаптивного ковзного двобічного експоненційного згладжування. Створено алгоритм для відновлення пропущених значень часових рядів з використанням розробленого методу. Застосовано метод адаптивного ковзного двобічного експоненційного згладжування для відновлення істинних закономірностей чисел Вольфа та даних про місячні значення сонячних радіофлюксів при довжині хвилі 10,7 см за даними Національного управління океанічних і атмосферних досліджень Сполучених Штатів Америки, виконано порівняння результатів з результатами, отриманими за методом 13-місячного

ковзного середнього. За значенням критерію близькості створений метод показав перевагу перед методом 13-місячного ковзного середнього. За значеннями статистичних характеристик створений алгоритм з використанням методу адаптивного ковзного двобічного експоненційного згладжування показав перевагу перед методом експоненційного згладжування при відновленні пропущених значень індексу F10.7. Виконано прогнозування даних про місячні значення сонячних радіофлюксів при довжині хвилі 10,7 см за допомогою методу адаптивного ковзного двобічного експоненційного згладжування та порівняння з результатами, отриманими за експоненційним підходом. За значеннями прогнозних характеристик розроблений метод показав перевагу перед традиційним експоненційним підходом при прогнозуванні значень сонячних радіофлюксів при довжині хвилі 10,7 см.

У четвертому розділі дисертаційної роботи визначено принципи системної методології, які застосовані під час розробки системи підтримки прийняття рішень для організації, що займається аналізом та прогнозуванням часових рядів.

Розроблено структурну схему частини аналізу та прогнозування часових рядів цієї системи підтримки прийняття рішень, її наведено на рис. 2, де АКФ – автокореляційна функція, ЧАКФ – часткова автокореляційна функція, МНК – метод найменших квадратів.

Виконано аналіз функціонування та основ роботи організації, що займається аналізом та прогнозуванням часових рядів. В організації, що займається аналізом та прогнозуванням часових рядів, виділено відділи інформаційних технологій, аналітики, кадрів. Створено різні рівні доступу до системи підтримки прийняття рішень у залежності від відділів для співробітників цієї організації. Виконано проектування бази даних. Створено архітектуру системи підтримки прийняття рішень, на якій показано основні етапи роботи з цією системою, її наведено на рис. 3.

Виконано програмну реалізацію системи підтримки прийняття рішень мовою програмування Java, створено опис для її користувачів.

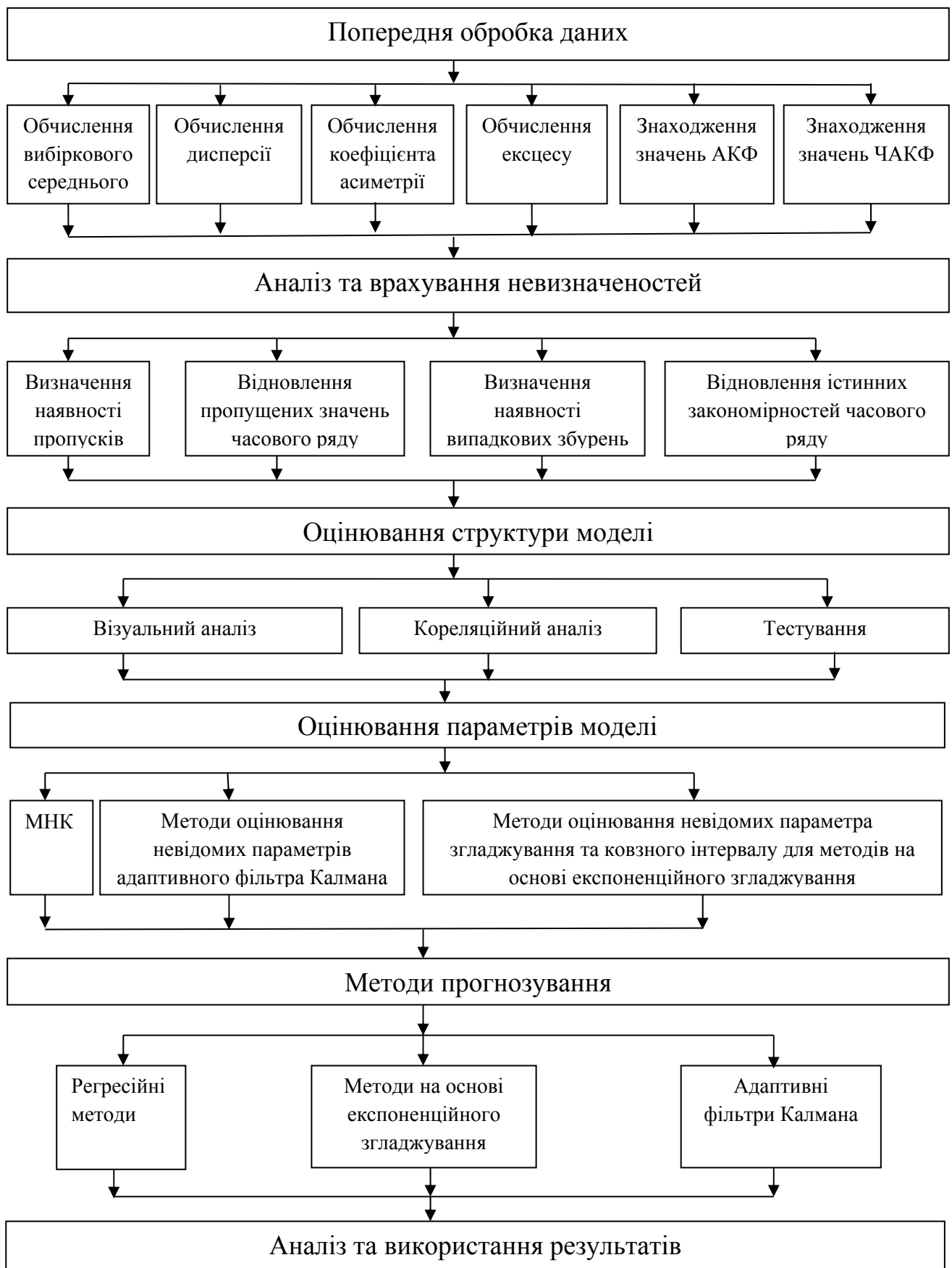


Рис. 2. Структурна схема частини аналізу та прогнозування часових рядів у системі підтримки прийняття рішень

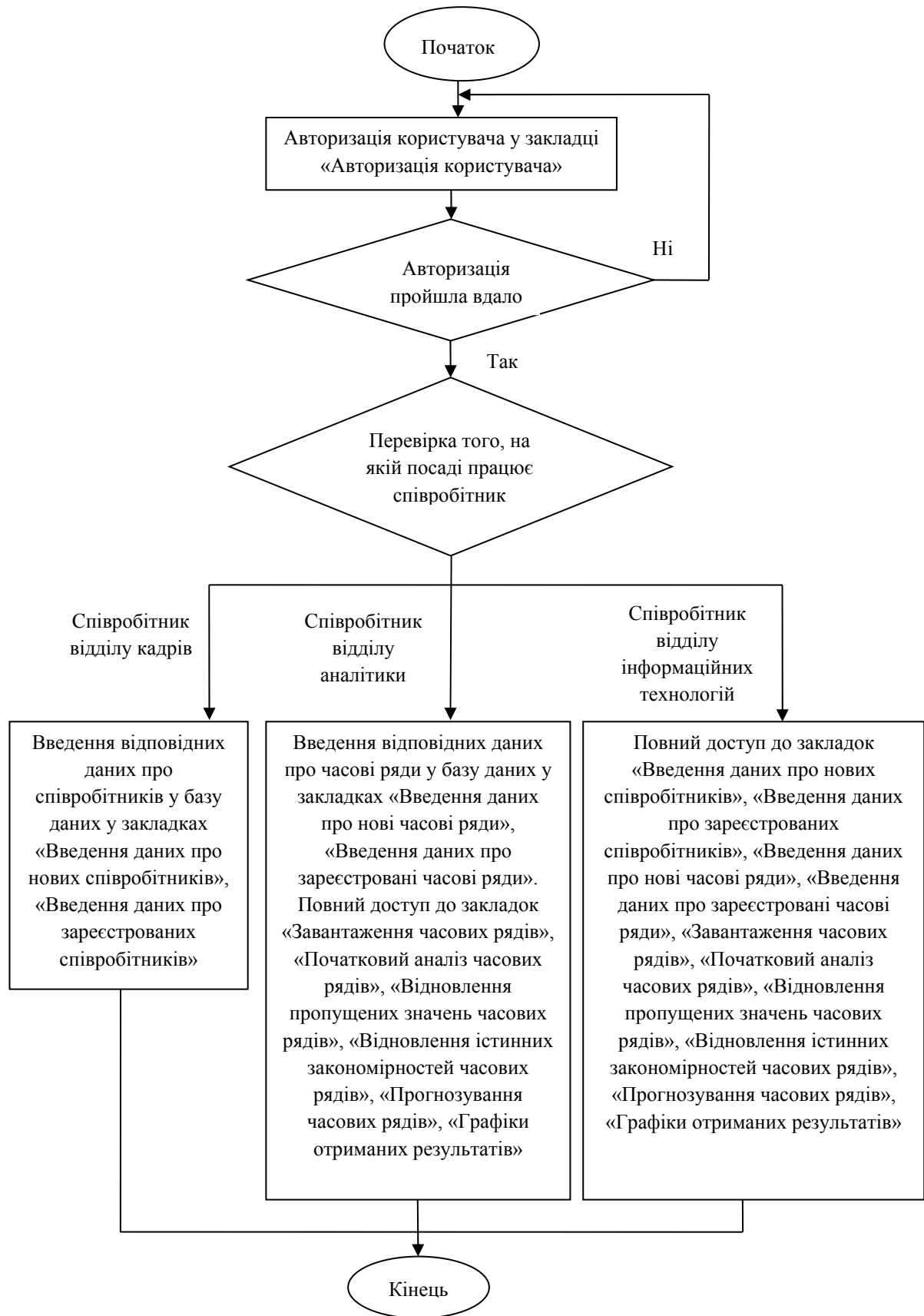


Рис. 3. Архітектура системи підтримки прийняття рішень

У створеній системі підтримки прийняття рішень реалізовано можливості виконання аналізу даних, традиційні методи, а також розроблені нові методи відновлення пропущених значень, відновлення істинних закономірностей та прогнозування часових рядів. Описано приклади її застосування для аналізу, відновлення пропущених значень, відновлення істинних закономірностей, прогнозування фінансово-економічних та сонячних даних.

ВИСНОВКИ

У дисертаційній роботі розв'язано науково-прикладну задачу стосовно розробки методів оцінювання параметрів математичних моделей, що змінюються за невідомим законом, та їх застосування для підвищення якості оцінок прогнозів розвитку часових рядів шляхом удосконалення системної методології прогнозування, розробки методів відновлення пропущених значень, істинних закономірностей та прогнозування часових рядів, розробки оригінального програмного продукту у формі системи підтримки прийняття рішень. За результатами виконаного дослідження можна зробити такі основні висновки:

1. Розроблено методи оцінювання математичного сподівання прискорення зміни значень вибірки даних з використанням повних та розріджених вибірок нев'язок субоптимального фільтра з пам'яттю два, та виконано порівняння створених методів з існуючими. Обрано кращі методи серед розроблених, та застосовано їх для розробки адаптивних фільтрів Калмана. Розроблені адаптивні фільтри Калмана показали перевагу перед традиційними методами за значеннями прогнозних характеристик під час прогнозування середньодобових цін Лондонської біржі металів на цинк ($MARE=4,1562\%$ для адаптивного фільтра Калмана № 1; $MARE=4,1176\%$ для адаптивного фільтра Калмана № 2).

2. Розроблено метод оцінювання математичного сподівання прискорення зміни значень вибірки даних, що змінюється за невідомим законом, з застосуванням процедури експоненційного згладжування до побудованого ряду однозалежних псевдовимірів цього параметру. Створено алгоритми прогнозування на основі фільтра Калмана з використанням розробленого методу оцінювання математичного сподівання прискорення зміни значень вибірки даних. Створені алгоритми прогнозування на основі фільтра Калмана показали перевагу перед традиційними методами за значеннями прогнозних характеристик під час прогнозування середньодобових цін на свинець за даними Лондонської біржі металів ($MARE=4,3465\%$ для алгоритму прогнозування № 1 на основі фільтра Калмана; $MARE=4,2352\%$ для алгоритму прогнозування № 2 на основі фільтра Калмана).

3. Розроблено метод двобічного експоненційного згладжування. Створено інтегральний критерій оцінки адекватності моделі. Описано застосування створеного методу для відновлення істинних закономірностей та прогнозування часових рядів. Розроблено алгоритм для відновлення пропущених значень часового ряду з використанням методу двобічного експоненційного згладжування. Метод двобічного експоненційного згладжування за значеннями статистичних

характеристик показав перевагу перед традиційними методами під час відновлення середньодобових значень цін на цинк за даними Лондонської біржі металів.

4. Розроблено метод відновлення пропущених значень для взаємозалежних часових рядів з використанням двобічного експоненційного згладжування. Створений метод показав перевагу перед традиційними методами за значеннями статистичних характеристик під час відновлення значень взаємозалежних показників сталого розвитку за даними Світового центру даних. Розроблено метод прогнозування для взаємозалежних часових рядів з використанням двобічного експоненційного згладжування. Створений метод показав перевагу перед методом експоненційного згладжування за значеннями прогнозних характеристик під час прогнозування показників сталого розвитку.

5. Розроблено метод ковзного двобічного експоненційного згладжування та алгоритм для відновлення пропущених значень ряду з використанням цього методу. Створений алгоритм показав перевагу перед методом експоненційного згладжування за значеннями статистичних характеристик під час відновлення даних про місячні значення сонячних радіофлюксів при довжині хвилі 10,7 см за даними Національного управління океанічних і атмосферних досліджень Сполучених Штатів Америки. Розроблено критерій близькості для оцінювання моделей під час відновлення істинних закономірностей часових рядів. Метод ковзного двобічного експоненційного згладжування показав перевагу перед методом 13-місячного ковзного середнього при відновленні істинних закономірностей індексу F10.7 за значенням критерію близькості. За значеннями прогнозних характеристик розроблений метод показав перевагу перед традиційним експоненційним підходом при прогнозуванні значень індексу F10.7.

6. Розроблено метод адаптивного ковзного двобічного експоненційного згладжування та алгоритм для відновлення пропущених значень часових рядів з використанням цього методу. За значенням критерію близькості створений метод показав перевагу перед методом 13-місячного ковзного середнього під час відновлення істинних закономірностей чисел Вольфа та даних про місячні значення сонячних радіофлюксів при довжині хвилі 10,7 см. За значеннями статистичних характеристик створений алгоритм з використанням методу адаптивного ковзного двобічного експоненційного згладжування показав перевагу перед методом експоненційного згладжування при відновленні пропущених значень індексу F10.7. За значеннями прогнозних характеристик метод адаптивного ковзного двобічного експоненційного згладжування показав перевагу перед традиційним експоненційним підходом при прогнозуванні місячних значень сонячних радіофлюксів при довжині хвилі 10,7 см.

7. Визначено принципи системної методології, які застосовані під час розробки системи підтримки прийняття рішень для організації, що займається аналізом та прогнозуванням часових рядів. Розроблено архітектуру і структурну схему частини аналізу та прогнозування часових рядів цієї системи підтримки прийняття рішень, створено різні рівні доступу до системи підтримки прийняття рішень в залежності від відділів для співробітників цієї організації. Виконано її програмну реалізацію мовою програмування Java, та створено опис для її

користувачів. У створеній системі підтримки прийняття рішень реалізовано можливості виконання аналізу даних, традиційні методи та розроблені нові методи відновлення пропущених значень, відновлення істинних закономірностей та прогнозування часових рядів. Наведено приклади її застосування для аналізу, відновлення пропущених значень, відновлення істинних закономірностей, прогнозування фінансово-економічних та сонячних даних.

СПИСОК ОПУБЛІКОВАНИХ ПРАЦЬ ЗА ТЕМОЮ ДИСЕРТАЦІЇ

1. Братусь О.В., Подладчиков В.М. Побудова багатовимірної моделі на основі фільтра Калмана й аналіз алгоритмів оцінювання її параметрів. *Наукові вісті НТУУ «КПІ»*. 2013. № 5 (91). С. 28–34. Особистий внесок: здобувачеві належать створені методи № 4–6 для оцінювання математичного сподівання прискорення зміни значень вибірки даних, використання розробленого методу № 5 для побудови адаптивного фільтра Калмана та подальшого прогнозування цін на цинк.
2. Братусь О.В., Подладчиков В.М. Ідентифікація змінних параметрів моделі для побудови алгоритму прогнозування. *Системні дослідження та інформаційні технології*. 2015. № 3. С. 131–141. Особистий внесок: здобувачеві належать метод оцінювання математичного сподівання прискорення зміни значень вибірки даних та створений алгоритм прогнозування на основі фільтра Калмана з використанням цього методу.
3. Братусь О.В., Подладчиков В.М., Бідюк П.І. Метод двобічного експоненційного згладжування для відновлення динамічних процесів. *Наукові вісті НТУУ «КПІ»*. 2016. № 6 (110). С. 15–21. Особистий внесок: здобувачеві належать розроблені метод двобічного експоненційного згладжування та інтегральний критерій адекватності моделі.
4. Братусь О.В. Система підтримки прийняття рішень з адаптивними блоками відновлення та прогнозування сонячних радіофлюксів. *Радіоелектроніка, інформатика, управління*. 2017. № 3. С. 36–43. (Роботу [4] опубліковано у журналі «Радіоелектроніка, інформатика, управління», що входить до переліку наукових фахових видань України та міжнародної наукометричної бази даних Web of Science).
5. Братусь О.В., Бідюк П.І., Болдак А.О. Розробка методів відновлення пропущених значень і прогнозування для взаємозалежних часових рядів. *Міжнародний науково-технічний журнал «Проблеми керування та інформатики»*. 2017. № 5. С. 13–21. (Роботу [5] опубліковано у журналі «Міжнародний науково-технічний журнал «Проблеми керування та інформатики»», що входить до переліку наукових фахових видань України, у повному обсязі перевидається англійською мовою під назвою «Journal of Automation and Information Sciences» та входить до міжнародної наукометричної бази даних Scopus). Особистий внесок: здобувачеві належать метод відновлення пропущених значень для взаємозалежних часових рядів з використанням двобічного експоненційного згладжування та метод прогнозування для взаємозалежних часових рядів з використанням двобічного експоненційного згладжування.

6. Bratus O. Development of adaptive moving two-sided exponential smoothing method for restoring and forecasting of time series. *EUREKA: Physics and Engineering*. 2017. No. 5. P. 13–21. (Роботу [6] опубліковано в іноземному виданні «EUREKA: Physics and Engineering»).
7. Братусь О.В., Подладчиков В.М. Розробка методів прогнозування на основі фільтра Калмана. *Системний аналіз та інформаційні технології*: матеріали 15-ї Міжнародної науково-технічної конференції SAIT 2013, м. Київ, ННК «ІПСА» НТУУ «КПІ», 27-31 травня 2013 р. Київ: ННК «ІПСА» НТУУ «КПІ», 2013. С. 251. Особистий внесок: здобувачеві належать методи оцінювання математичного сподівання прискорення зміни значень вибірки даних та адаптивний фільтр Калмана.
8. Братусь О.В., Подладчиков В.М. Пошук закономірностей динамічних процесів на основі двобічного експоненційного згладжування. *Контроль і управління в складних системах (КУСС-2014)*: тези доповідей XII міжнародної конференції, м. Вінниця, 14-16 жовтня 2014 р. Вінниця: ВНТУ, 2014. С. 28. Особистий внесок: здобувачеві належить метод двобічного експоненційного згладжування.
9. Братусь О.В. Система підтримки прийняття рішень для аналізу та прогнозування цін на метали. *Інтелектуальні системи прийняття рішень і проблеми обчислювального інтелекту*: матеріали міжнародної наукової конференції, с. Залізний Порт, 24-28 травня 2016 р. Херсон: ПП Вишемирський В.С., 2016. С. 185–186.
10. Братусь О.В. Ідентифікація змінної дисперсії шуму вимірів моделі у задачах прогнозування. *Математика. Інформаційні технології. Освіта*: тези доповідей V міжнародної конференції, м. Луцьк - с. Світязь, 5-7 червня 2016 р. Луцьк: ПП Іванюк В.П., 2016. С. 58–59.
11. Братусь О.В., Подладчиков В.М., Подладчикова Т.В. Метод ковзного двобічного експоненційного згладжування для відновлення закономірностей динамічних процесів. *Математичне та імітаційне моделювання систем. МОДС 2016*: тези доповідей одинадцятої міжнародної науково-практичної конференції, с. Жукин, 27 червня - 1 липня 2016 р. Чернігів: ЧНТУ, 2016. С. 28–31. Особистий внесок: здобувачеві належать метод ковзного двобічного експоненційного згладжування та критерій близькості.

АНОТАЦІЯ

Братусь О.В. Методи прогнозування нестационарних часових рядів на основі двобічного експоненційного згладжування та оптимальної фільтрації. – На правах рукопису.

Дисертація на здобуття наукового ступеня кандидата технічних наук за спеціальністю 01.05.04 – системний аналіз і теорія оптимальних рішень. – Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», Міністерство освіти і науки України, Київ, 2019.

Розроблено методи оцінювання математичного сподівання прискорення зміни значень вибірки даних, їх застосовано для розробки адаптивних фільтрів Калмана та

алгоритмів прогнозування на основі фільтра Калмана. Для відновлення пропущених значень, істинних закономірностей та прогнозування часових рядів розроблено методи двобічного експоненційного згладжування, ковзного двобічного експоненційного згладжування та адаптивного ковзного двобічного експоненційного згладжування. Розроблено методи відновлення пропущених значень та прогнозування для взаємозалежних часових рядів з використанням двобічного експоненційного згладжування. Розроблено інтегральний критерій адекватності моделі та критерій близькості. Створено систему підтримки прийняття рішень, в якій реалізовано традиційні та розроблені методи.

Ключові слова: адаптивний фільтр Калмана, метод двобічного експоненційного згладжування, взаємозалежні часові ряди, інтегральний критерій адекватності моделі, критерій близькості, метод ковзного двобічного експоненційного згладжування, метод адаптивного ковзного двобічного експоненційного згладжування, система підтримки прийняття рішень, системний аналіз.

АННОТАЦІЯ

Братусь Е.В. Методы прогнозирования нестационарных временных рядов на основе двустороннего экспоненциального сглаживания и оптимальной фильтрации. – На правах рукописи.

Диссертация на соискание ученой степени кандидата технических наук по специальности 01.05.04 – системный анализ и теория оптимальных решений. – Национальный технический университет Украины «Киевский политехнический институт имени Игоря Сикорского», Министерство образования и науки Украины, Киев, 2019.

Разработаны методы оценивания математического ожидания ускорения изменения значений выборки данных, они применены для разработки адаптивных фильтров Калмана и алгоритмов прогнозирования на основе фильтра Калмана. Для восстановления пропущенных значений, истинных закономерностей и прогнозирования временных рядов разработаны методы двустороннего экспоненциального сглаживания, скользящего двустороннего экспоненциального сглаживания и адаптивного скользящего двустороннего экспоненциального сглаживания. Разработаны методы восстановления пропущенных значений и прогнозирования для взаимозависимых временных рядов с использованием двустороннего экспоненциального сглаживания. Разработаны интегральный критерий адекватности модели и критерий близости. Создана система поддержки принятия решений, в которой реализованы традиционные и разработанные методы.

Ключевые слова: адаптивный фильтр Калмана, метод двустороннего экспоненциального сглаживания, взаимозависимые временные ряды, интегральный критерий адекватности модели, критерий близости, метод скользящего двустороннего экспоненциального сглаживания, метод адаптивного скользящего двустороннего экспоненциального сглаживания, система поддержки принятия решений, системный анализ.

ABSTRACT

Bratus O.V. Methods of forecasting of non-stationary time series based on two-sided exponential smoothing and optimal filtering. – On the rights of manuscript.

Dissertation for scientific degree of candidate of technical sciences on the speciality 01.05.04 – system analysis and optimal decisions theory. – National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Ministry of education and science of Ukraine, Kyiv, 2019.

The dissertation work is dedicated to development and application of applied scientific methodology of system analysis for solving actual tasks of restoring of missing values of time series, restoring true regularities of researched processes, forecasting of time series and performing automated processing of time series using created decision making support system.

Methods of estimation of mathematical expectation of acceleration of values change of data samples using full and rarified samples of residuals of suboptimal filter with memory two were developed in this work. The best methods were selected between developed methods, they were used for development of adaptive Kalman filters. Adaptive Kalman filters showed better results in comparison to traditional methods in forecasting of daily average prices of zinc by the London metal exchange data.

Method of estimation of mathematical expectation of acceleration of changes of time series values, which is changed based on the unknown law, using the exponential smoothing procedure to the constructed series of one-dependent pseudo-measurements of this parameter was developed. Forecasting algorithms based on Kalman filter were created using developed method. Created forecasting algorithms showed superiority in comparison to traditional methods by forecasting characteristics in forecasting of daily average prices of lead.

Two-sided exponential smoothing method and algorithm for restoring of missing values of time series using this method were developed. The integral criterion of model adequacy was created. Application of the created method for restoring of true regularities and forecasting of time series was described. Created method showed superiority in comparison to traditional methods by statistical characteristics in restoring of daily average prices of zinc.

Method for restoring missing values for mutually dependent time series using two-sided exponential smoothing was developed. This method showed superiority in comparison to traditional methods by statistical characteristics in restoring mutually dependent indices of sustainable development by the data of World data center. Method for forecasting of mutually dependent time series using two-sided exponential smoothing method was developed. This method showed superiority in comparison to exponential smoothing method in forecasting of sustainable development indices.

Moving two-sided exponential smoothing method and algorithm for restoring of missing values of time series using this method were developed. This algorithm showed superiority in comparison to exponential smoothing method by statistical characteristics in restoring of monthly values of solar radio fluxes at a wavelength of 10.7 cm. The proximity criterion for estimation of models in restoring of the true regularities of time

series evolution was developed. Moving two-sided exponential smoothing method showed better results in comparison to traditional methods in restoring of the true regularities and forecasting of solar data.

Adaptive moving two-sided exponential smoothing method and algorithm for restoring of missing values of time series using this method were developed. Created method showed superiority in comparison to traditional methods in the restoring of missing values, true regularities and forecasting of solar data.

Principles of system methodology were determined, which were used during development of decision making support system for organization, which is working on analysis and forecasting of time series. Architecture and structural scheme of part of analysis and forecasting of decision making support system were developed, different levels of access to this system were created according to departments for employees of this organization. Software implementation of decision making support system was performed, and description for its users was created. Opportunities of data analysis performing, traditional methods and created new methods of restoring of missing values, restoring of true regularities and forecasting of time series were implemented in the decision making support system. Examples of its using for work with financial-economic and solar data were provided.

Key words: adaptive Kalman filter, two-sided exponential smoothing method, mutually dependent time series, integral criterion of model adequacy, proximity criterion, moving two-sided exponential smoothing method, adaptive moving two-sided exponential smoothing method, decision making support system, system analysis.